图书情报工作
第 62 卷 第 22 期　2018 年 11 月

科技评价指标值与评价属性背离及修正研究*

——基于多属性评价视角

周娟美[1]　郭强华[2]　王作功[3]　俞立平[4]

[1] 中北大学经济与管理学院　太原 030051　[2] 云南大学经济学院　昆明 650504

[3] 贵州财经大学中国西部绿色发展战略研究院　贵阳 550025

[4] 浙江工商大学管理工程与电子商务学院　杭州 310018

摘要：［目的/意义］分析科技评价中指标区分度异常、数据分布有偏的深层次原因，认为本质上这是评价指标值与评价属性的背离现象，即评价指标值不能较好地体现评价属性的本质含义。［方法/过程］提出一种新的降低评价指标值与评价属性背离的方法——对数中位数标准化，并以 JCR2016 数学期刊为例进行实证分析。［结果/结论］研究结果表明：引文指标更容易出现评价指标值与评价属性背离问题；可以从多角度判定评价指标值与评价属性背离问题，如指标内涵分析、及格率、离散系数、中位数极大值比、集中度指数 HHI 等；采用对数中位数标准化可以大幅降低评价指标值与评价属性背离问题；建议评价中如出现指标值与属性背离，采用对数中位数处理后的数据进行评价。

关键词：区分度　数据分布　科技评价　影响力指标　对数中位数标准化

分类号：G302

DOI：10.13266/j.issn.0252-3116.2018.22.012

1 引言

随着创新型国家建设步伐加快以及双一流大学建设的推进，科技评价得到了越来越多的重视，对学术评价自身的要求也越来越高。政府部门、学术界、公众也展开了越来越多的关于科技评价的讨论，涉及评价机制、评价目的、评价技术、评价应用等诸多方面。做好评价工作的每一个细节，尤其是解决好一些系统性、普遍性的问题，才能提高评价的公信力和科学性，做到公平、公正、公开。

一些文献计量指标的数据分布有偏、区分度异常是大家公认的，或者说，文献计量指标的数据分布本该如此。以 JCR2016 数学期刊为例，总共 310 种期刊中，总被引频次最高的期刊为 19 970 次，超过 10 000 的只有 5 种期刊，而低于 800 的有 165 种期刊，超过期刊数量的一半，总被引频次低分区域数据密集，区分度很低，而高分区期刊数量很少，区分度又过高。从另外一个角度看，假设将最高分的 60% 视为及格线，那么 310 种期刊中，只有 3 种期刊及格，无论对于学者还是期刊来说，总被引频次所反映的期刊影响力不应该有如此巨大的差距。至于特征因子，低分区数据更加密集，许多期刊精确到小数点后 5 位也不能有效区分。如果在论文质量层面出现较大差距也许是正常的，但无论是影响因子还是特征因子，都是反映学术期刊影响力的指标，并不代表真实的论文质量。那些重大科技创新的质量差距大是正常的，而影响力差距这么大是值得商榷的。

公众对单一指标数据分布的要求与综合评价结果的要求是不一致的。对于单一指标，由于其专业性，数据分布不服从正态分布是正常的；对于综合指标，除了图书馆情报与文献学、统计学等领域的少数专业人士外，绝大多数公众和大多数作者等一般是难以理解的，

* 本文系山西省高等学校人文社会科学重点研究基地项目“山西省科技金融服务平台与支持政策研究”（项目编号：2017335）和贵州省科技厅软科学研究项目“贵州承接台湾产业转移研究——基于产业园视角”（项目编号：黔科合基础［2016］1510-2）研究成果之一。

作者简介：周娟美（ORCID：0000-0002-9596-5918），副教授，博士；郭强华（ORCID：0000-0001-7882-0218），教授，博士；王作功（ORCID：0000-0002-5835-3883），教授，博士；俞立平（ORCID：0000-0001-9079-1165），教授，博士，博士生导师，通信作者，E-mail：yvliping@126.com。

收稿日期：2018-06-03　修回日期：2018-07-14　本文起止页码：100-108　本文责任编辑：刘远颖

比如认为正常情况下期刊影响力应该服从正态分布，呈现“中间多两头少”的规律，如果某学科期刊综合影响力只有5%的期刊大于60分，往往很难理解。这里影响力就是评价属性，所谓评价属性，就是单一或若干评价指标试图说明的问题或评价的目的。

学术期刊影响力的差距有影响因子、特征因子指标所反映的那么大吗？影响因子、特征因子等指标能否很好地代表学术期刊的影响力属性？一些文献计量指标数据分布有偏与区分度异常的背后原因是什么？期刊评价指标数据分布偏倚对综合评价结果有什么影响？对相关问题开展研究，探索其中隐含的深层次原因，分析改善区分度的方法以及对评价的影响，改进评价指标的数据分布使之更符合评价属性特征与公众认知，不仅有利于推进评价理论，而且对于做好科技评价工作，提高评价质量具有重要意义。

许多学者发现学术期刊评价指标数据有偏，P. Vinkler[1]证明了引文分布的右偏性，认为发表在影响因子较高期刊上的论文仅仅是提供了获得高被引的可能性，并不能合理地衡量期刊的影响力，影响因子对期刊真实影响力的衡量存在较大偏差。P. O. Seglen[2]认为引文分析数据分布是典型的偏态分布，不服从正态分布，具有幂律分布特征。L. Bornman 等[3]认为在一些研究中用平均值来反映引文的集中趋势是不对的，因为引文并不服从正态分布。R. Adler 等[4]认为，引文均值反映的更多是高被引论文的引用值，根据幂律法则，由于引用数据的分布通常是右偏态分布的。A. F. J. Raan [5]用平均值来反映引文数据的做法使得评价结果远远大于或小于国际引文影响标准。

科技评价区分度问题是评价技术中的一个小问题，往往关注不够。区分度又叫辨识度或粒度，是指统计指标在区分各评价单位某方面特征时的能力与效果，反映了功效得分分布的分散程度[6]。P. M. Davis[7]发现对于总体影响力比较低的期刊群来说，连续等级期刊之间的特征因子值差别很小，离散程度很低。JCR 又推出了标准特征因子指标，就是为了改善特征因子区分度较低的不足。其实影响因子百分位指标也是为了提高区分度而设计的，它主要是基于排序变换，是非参数数据。h 指数[8]虽然是个设计精巧的评价指标，但 W. Glänzel[9]认为 h 指数缺乏区分度，大部分科研人员的 h 指数处于相近水平，难以对个体绩效做出有效评估。俞立平等[10]分析了高校期刊分级中存在的区分度低、综合性期刊和新兴学科交叉学科重视不够、本校学报定级偏高、期刊分级周期过长等问题。文东茅等[11]认为等级赋分降低了区分度，在一定程度上会影响学科特别拔尖者的相对优势。

提高区分度的方法首先是从评价指标设计角度开始，在文献计量学中以提高区分度为主的改进主要是关于 h 指数方面的研究。F. Ruane 等[12]提出 h_{rat} 指数，解决了相同 h 指数科学家的可比问题。叶鹰[13]把 f 指数放大 100 倍取对数后再放大 100 倍定义为对数 f 指数 fl，在评价上使区分度更细化。付鑫金等[14]发现网络 h 指数的区分度不够，配合使用网络 g 指数及学术差、学术势来共同评价高校的网络学术影响力，效果更好。许新军[15]认为 h_d 指数在区分度、时效性、考察范围、稳健性等方面明显优于 h_c 指数，而在严谨性方面 h_d 指数则略逊于 h_c 指数。

关于多属性评价中提高区分度的方法，刘学之、杨泽宇等[16]提出在处理非均匀分布的指标数据集合时，局部集中分布数据无法有效地划分层级，缺乏辨识性，采用 Logistic 曲线函数的特性构建 S 型曲线模型，可对指标数据进行非线性标准化处理。郭亚军等[17]认为综合评价结果不仅受到指标权重的影响，很大程度上也取决于指标标准化的方法，并基于尽可能反映多个被评价对象之间局部和整体差异的原则，提出了拉开档次法。俞立平等[18]认为标准 TOPSIS 是一种对较好期刊区分度较好、对弱势期刊区分度较差的评价方法。俞欣辰、潘有能[19]研究发现，多属性评价会显著降低期刊评价指标的数据偏倚水平，TOPSIS 法的效果最好，熵权法次之，而因子分析法最差，更加接近原始评价指标的区分度和分布状态。

从现有研究看，文献计量指标的数据分布有偏现象已经引起广泛的重视，只有少数学者反思其中的原因。评价的区分度问题属于评价的基础问题，存在于一切评价中，不是科技评价中独有。在科技评价中，h 指数由于区分度较低得到了更多的关注，但是在多属性评价中没有得到应有的重视。影响区分度的因素较多，涉及到指标设计、指标评分制、数据标准化方法、评价方法等。总体上，关于多属性评价数据分布与区分度问题的系统性研究不够，在以下几个方面有待深入：①评价指标是用来反映科技评价中某个属性的，这种属性如果出现数据扭曲，导致评价对象的属性差距不符合实际，数据分布和区分度异常，深层次的原因是什么？②为了改进评价指标数据分布与区分度的问题，应该采用什么方法对数据进行处理？③线性评价方法与非线性评价方法对评价结果的数据分布与区分度会产生何种影响？④对评价区分度如何测度，在现有方

法的基础上没有更好的测度方法。

本文以 JCR2016 数学学术期刊为例，主要从评价指标与多属性评价方法角度进行分析，因此不讨论评价指标设计层面的区分度问题，重点探讨评价指标原始数据能否客观反映评价对象的某种属性差距，分析其中深层次的原因，并从指标数据标准化方法角度探讨改进数据分布与区分度的方法以及对多属性评价结果数据分布与区分度的影响。

2 评价区分度的理论分析

2.1 评价指标数据分布与区分度异常原因

由于相关问题比较复杂，本文重点从学术期刊评价角度进行分析，探讨许多学术期刊评价指标数据分布有偏、区分度异常背后深层次的原因。

2.1.1 评价指标数据分布与区分度问题一定程度上被掩盖　公众往往难以接受科技评价结果的严重偏态现象，比如评价结果及格率在 5% 以下。常见的科技评价比如大学评价、产业园评价、学术期刊评价、科技绩效评价、学科评价等，很多时候并不公布评价得分，只有排名，这样就无法看到评价对象之间的差距，因此一定程度上也掩盖了这个问题。

2.1.2 只有极少数科技评价指标呈现偏态才是正常的　在科技评价中，大多数指标应该服从或者接近正态分布，极少数评价指标呈现偏态分布也是正常的。比如创新水平，那些真正的原始创新与重大创新，如相对论、量子理论等，这些反映创新能力的指标分布呈现偏态分布，区分度异常，公众是可以接受的。问题是真正反映创新水平的指标极其缺乏，原创水平一般只有通过同行评议才能评价，难以通过文献计量指标进行评价，甚至有时同行评议也难以做出客观评价。因此通常情况下，不宜有太多评价指标出现数据分布与区分度问题，如果出现，一定要认真分析该指标的本质含义。

2.1.3 现有文献计量指标难以衡量创新　在文献计量影响力指标中，如此众多的指标呈现偏态分布是一种不正常现象，因为这些指标难以衡量创新。在学术期刊评价中，相关指标主要分为三大类：①影响力指标，反映了期刊论文被引用的相关情况，如影响因子、5 年影响因子、他引影响因子等；②时效性指标，如被引半衰期、引用半衰期等；③来源指标，如基金论文比、地区分布数、平均作者数等，当然其他还有可能有一些编辑出版指标等。学术期刊中最重要的指标论文质量，或者说论文的创新水平，由于难以获得相关数据，往往是无法直接评价的，这方面的指标非常少。当然，学术期刊影响力、时效性、来源指标往往与期刊的学术质量呈现正相关，这就是从这些方面研究的意义所在。

2.1.4 文献计量评价指标值与评价属性背离　根据以上分析，只有原创性的成果才应该服从偏态分布，出现较大的区分度异常，而现有文献计量指标根本无法评价原创性成果，但是其数据分布又表现为偏态分布，那么原因就只有一个，本文将其称为“评价指标值与评价属性的背离”，即文献计量指标原始数据包含的信息存在扭曲现象，没有反映评价指标的真实属性，表现为数据分布有偏、区分度异常。期刊影响力、期刊时效性、期刊来源指标原始数据并没有反映真实属性信息。总被引频次相差悬殊并不代表期刊影响力属性相差悬殊，或者说，原始数据将期刊影响力属性差距拉大了，总被引频次 16 000 的期刊，其影响力并不是总被引频次 800 期刊的 20 倍。类似的指标还有许多，5 年影响因子、特征因子等都有类似问题。抛开评价指标，期刊影响力的真实水平其实应该接近正态分布才更加合理，注意影响力真实水平与衡量影响力的指标值是两个概念，由于衡量影响力的指标原始数据信息扭曲，才使得人们误认为期刊影响力属性真实水平也是扭曲的，这才是问题的关键。

2.2 评价指标的数据分布与区分度会影响综合评价结果

以上是从评价指标角度分析数据分布与区分度问题，下面进一步从多属性评价角度进行分析。

多属性评价是基于文献计量指标，通过一定的方式赋权，选择某种评价方法进行评价后的结果。现有的几种学术期刊主流评价，如北京大学中文核心期刊、中国科学技术信息研究所 CSTPCD、中国科学院文献情报中心 CSCD、中国社会科学院核心期刊等，虽然评价体系不同、评价方法不同、赋权不同，但评价原理都大同小异。

评价指标的区分度与数据分布必然会影响评价结果，所以评价结果也有数据分布和区分度问题。俞立平、刘爱军[20]研究发现，指标数据偏倚会影响期刊一般水平的判断，指标数据右偏会导致期刊评价值偏低，最好选取数据偏倚情况相对较好的指标来评价期刊平均水平。

2.3 改进评价指标数据分布与区分度的思路

对原始指标进行非线性化处理，使得处理后的数据更加能表达其属性含义，并且改善数据的区分度，使得其更加接近正态分布。指标属性与无量纲化处理方

法之间的关系密不可分，应根据原始数据的相关特性设计指标无量纲化方法[21]。应该针对不同评价指标的数据分布与区分度特点，选择相应的非线性化处理方法。

本文拟对原始文献计量指标取自然对数来改善其区分度与数据分布。联合国开发计划署在计算人口发展指数时，对于国民收入指标，为了体现每增加一美元收入提升人类发展水平的边际效用递减，先作自然对数处理，而后使用线性无量纲化方法得到国民收入分指数[22]。苏为华[23]认为对数处理有很好的性质，包括严格单调，具有较强的凹性，从而能改变分布形态，能缩减值域区间，因此对增长较快、跨度大的指标处理效果好。封婷[24]认为观测值集中在低水平的情况并不少见，呈现右偏分布，使截面研究中指标取值集中在低水平，在这种情况下，使用非线性功效函数应为凹函数。

2.4 研究设计

本文研究设计的技术路线如图1所示。第一种处理方法是对原始评价指标不作任何变换，进行评价。第二种处理方法是通过取对数与新的标准化方法综合采用，适当改变原始指标的数据分布和区分度，然后再进行评价。最后比较这两种处理方法数据分布与区分度的区别，并得出研究结论。

2.4.1 原始指标的数据分布及区分度的判断　首先要对文献计量指标的含义进行分析，该指标属于什么性质的指标，如果该指标体现了创新水平，可能偏态分布是正常的，没有必要对评价指标数据进行进一步处理。如果该指标就是影响力、时效性或者一般来源指标，那么就要对这些指标原始数据分布进行判断，主要进行正态分布检验，常用的方法是 Jarque-Bera 检验。

关于区分度的判断方法很多，王连芬、张少杰[25]在产业竞争力测度中，提出采用离散系数表示区分度。注意要使得数据分布更加均匀，离散系数是一个反向指标。俞立平、刘爱军[20]提出采用基尼系数的原理测度区分度。由于基尼系数计算又过于复杂，因此本文借用 A. O. Hirschman[26]提出的郝氏指数 HHI 来测度区分度，其计算公式如下：

$$HHI=\sum_{i=1}^{m}\left(\frac{x_{ij}}{\sum_{i=1}^{m}x_{ij}}\right)^{2} \qquad \text{公式(1)}$$

公式(1)中，x_{ij}为评价指标，i 为评价对象序号，j 为评价指标序号，m 为评价对象数量，HHI 为郝氏指数，即评价指标份额的平方和，该指标是个反向指标，越大说明评价指标分布越不均衡，集中在少数期刊中。

此外，为了对评价指标的区分度进行分析，还增加了及格数和及格率指标，即评价值达到极大值60%的评价对象的数量和比例。另外还采用中位数极大值比指标，以反映中位数的相对位置。

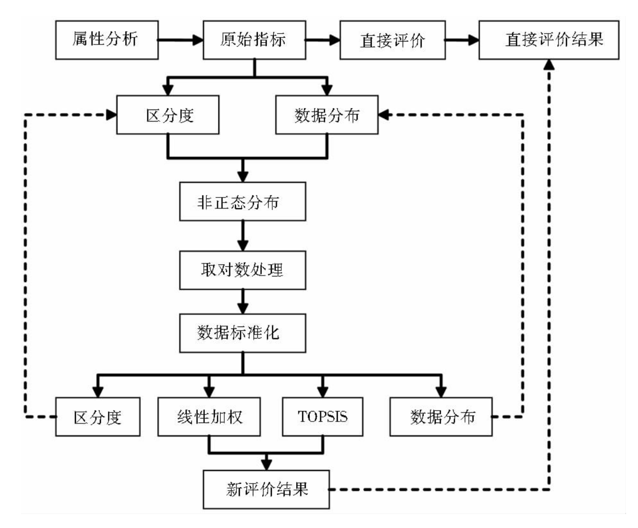


图1　研究设计

2.4.2 原始指标取自然对数处理　对于需要取自然对数的评价指标，全部取自然对数，需要注意的是，少数指标值为0，无法取对数，必要时需要做一下数据处理，比如将所有评价指标增加极大值的1%。

2.4.3 明确评价目的对区分度的隐性要求　在科技评价中，评价目的对区分度的要求往往是隐性的，但是也有一些基本要求，比如评价结果大于60分，或者经过标准化后大于60分。这就要求评价指标最好也能具备这样的性质，一个简单的判断方法是看中位数的值，最好大于极大值的一半。如果达不到这个要求，在数据标准化时就要进行必要的调整。

2.4.4 对数中位数标准化　评价指标数据标准化的目的，是通过线性变换，将原来已经取对数的评价指标中位数值进行适当提高，比如达到极大值的一半，以提高公众对评价指标值的接受能力。数据处理分为两种情况，第一种情况是，对于中位数排在极差一半位置以上的指标采用传统标准化方法进行处理，因为在这种情况下，已经比较接近公众的认知了；第二种情况是，对于中位数排在极差一半位置后的指标，则增加该差值，使得标准化后极小值等于该值，避免标准化后的指标在低水平区域密集。本文将这种标准化方法称为“对数中位数标准化”方法。令中位数与极差位置的相对差距为 D_{ij}，则：

$$D_{ij}=\frac{0.5\{\max[\ln(x_{ij})]-\min[\ln(x_{ij})]\}-median[\ln(x_{ij})]}{\max[\ln(x_{ij})]-\min[\ln(x_{ij})]}\times 100 \quad 公式(2)$$

对于第二种情况，采用对数中位数标准化方法进行标准化：

$$X_{ij}=D_{ij}+(100-D_{ij})\times\frac{\ln(x_{ij})-\min[\ln(x_{ij})]}{\max[\ln(x_{ij})]-\min[\ln(x_{ij})]} \quad 公式(3)$$

对于第一种情况，从增加区分度的角度，传统标准化方法标准化：

$$X_{ij}=\frac{\ln(x_{ij})-\min[\ln(x_{ij})]}{\max[\ln(x_{ij})]-\min[\ln(x_{ij})]}\times 100 \quad 公式(4)$$

2.4.5 多属性评价及分析 基于以上分析，笔者选择多属性评价方法进行评价，以 TOPSIS 评价方法为例，对原始数据取对数前后进行比较，并比较评价结果，从而进一步说明指标区分度、数据分布处理对评价的影响。

3 研究数据与实证结果

3.1 研究数据

本文以 JCR2016 数学期刊为例进行研究，JCR2016 公布的评价指标主要有 11 个，分别是：总被引频次、他引影响因子、影响因子、即年指标、影响因子百分位、5 年影响因子、特征因子、标准化特征因子、论文影响分值、被引半衰期、引用半衰期。由于影响因子百分位是一种排序转换，具有非参数性质，因此该指标不选取。此外被引半衰期与引用半衰期两个指标，凡是时间超过 10 年的就没有公布具体数据，只简单显示大于 10，故也没有选取。因此实际选取的指标均是期刊影响力指标。

JCR2016 数学期刊共有 310 种，由于部分年度少数期刊存在数据缺失，故将这些期刊删除，经过整理后还有 294 种期刊。原始指标数据基本描述统计如表 1 所示：

表 1 原始数据描述统计

指标	均值	中位数	极大值	极小值	标准差
总被引频次	1 448.136	689.000	19 970.000	104.000	2 290.007
影响因子	0.788	0.650	4.692	0.221	0.568
他引影响因子	0.725	0.579	4.635	0.124	0.561
5 年影响因子	0.864	0.727	4.105	0.237	0.601
特征因子	0.004 63	0.002 35	0.049 84	0.000 20	0.006 64
标准特征因子	0.531	0.269	5.712	0.023	0.761
论文影响分值	0.913	0.641	6.615	0.077	0.963
即年指标	0.188	0.129	2.500	0.000	0.219

3.2 原始指标的数据分布与区分度情况

指标数据分布与区分度如表 2 所示。中位数极大值比反映了中位数所处的位置，该值最大的是 5 年影响因子，为 0.177，也就是说满分 100 分的话，有一半期刊的得分在 17.7 分以下，说明指标值较低时数据拥挤情况严重，区分度差。离散系数中，大部分位于 2 倍标准差以内，离散系数小于 2.0，只有引用半衰期离散系数较大。从郝氏指数 HHI 看，总体上均较低，由于郝氏指数最早是用于衡量企业垄断水平的，因此绝对值水平高低有待进一步分析。

表 2 原始指标数据分布与区分度

指标	中位数/极大值	离散系数	HHI 系数	及格期刊种数	偏度 S	峰度 K	JB 检验	概率 p
总被引频次	0.035	1.581	0.011 9	3	4.016	23.612	5 994.545	0.000
影响因子	0.138	0.721	0.005 2	7	3.243	17.460	3 076.852	0.000
他引影响因子	0.125	0.773	0.005 4	5	3.308	17.983	3 286.246	0.000
5 年影响因子	0.177	0.696	0.005 0	8	2.973	14.145	1 954.897	0.000
特征因子	0.047	1.433	0.010 4	5	3.668	19.406	3 956.367	0.000
标准特征因子	0.047	1.433	0.010 4	5	3.668	19.406	3 956.613	0.000
论文影响分值	0.097	1.056	0.007 2	6	3.484	18.406	3 502.304	0.000
即年指标	0.052	1.164	0.008 0	1	5.200	47.209	25 267.110	0.000
均值	0.090	1.107	0.008	5	3.695	22.203	6 374.367	0.000

如果每个指标的极大值为 100 分，60 分作为及格线，那么 294 种期刊中，8 个指标中，及格最多的期刊也只有 8 个，绝大多数期刊评价指标得分均为“不及格”水平，也就是说，评价指标数据对期刊的实际影响力呈现一定的扭曲，难以得到多数人认同。

从数据分布看，所有指标均不服从正态分布，均拒绝了 Jarque-Bera 检验的原假设。所有指标的偏度大于 0，呈现右偏特征。

3.3 数据处理后的结果分布

首先对 8 个指标全部取对数，并对 8 个指标的中位数值进行分析，发现即年指标的中位数为 –2.048，而极大值与极小值之和的一般位置为 –2.996，即中位

数位于前部,这是唯一的一个指标,因此采用常规方法即公式(4)进行标准化,其他 7 个指标采用公式(3)进行标准化。数据经处理后,数据分布与区分度如表 3 所示:

表 3　数据处理后数据分布与区分度

指标	中位数/极大值	离散系数	HHI 系数	及格期刊	偏度 S	峰度 K	JB 检验	概率 p
总被引频次	0.448	0.370	0.003 9	63	0.540	2.960	14.293	0.001
影响因子	0.449	0.330	0.003 8	44	0.612	3.795	26.102	0.000
他引影响因子	0.468	0.310	0.003 7	53	0.498	3.797	19.924	0.000
5 年影响因子	0.459	0.351	0.003 8	50	0.718	3.956	36.460	0.000
特征因子	0.476	0.345	0.003 8	69	0.413	3.124	8.534	0.014
标准特征因子	0.475	0.346	0.003 8	69	0.413	3.123	8.556	0.014
论文影响分值	0.489	0.355	0.003 8	76	0.240	3.182	3.217	0.200
即年指标	0.621	0.293	0.003 7	179	-1.960	7.946	487.905	0.000
均值	0.486	0.337	0.003 8	75.38	0.184	3.985	75.624	0.029

原始数据指标的中位数极大值比的平均值为 0.090,50% 的数据处在 9% 的最低水平,经过取对数标准化处理后,50% 的数据处在 48.6% 的水平,已经非常均匀了。从离散系数看,原始数据的离散系数均值为 1.107,经过处理后离散系数的均值为 0.337,离散系数降低的原因本质上是由于数据分布更加均匀。从郝氏指数 HHI 看,原始数据的 HHI 指数的均值为 0.008,经过处理后的 HHI 指数为 0.003 8,得到大幅度的降低,因为该指标是反向指标,说明数据的区分度更好,分布较为均匀。

从评价指标的及格率看,8 个指标平均及格的期刊数为 75.38 种,相对于数据处理前及格期刊的平均值只有 5 种期刊,已经得到极大改善。

从数据分布看,数据处理后,论文影响分值服从正态分布,其他指标虽然不服从正态分布,但是更加接近正态分布,原始数据的 Jarque-Bera 检验均值为 6 374.367,经过处理后 Jarque-Bera 检验值为 75.624,得到了大幅降低。

3.4　评价结果的比较

为了比较数据处理前后评价结果的区分度和数据分布特征,本文同时采用线性加权汇总与 TOPSIS 进行评价和比较分析。简洁起见,本文采用等权重法,评价结果的部分数据如表 4 所示(前 30 种)。即使采用相同的评价方法,评价结果排序还是有所差异。

表 4　数据处理前后部分评价结果

期刊名称	线性加权处理前	排序	线性加权处理后	排序	TOPSIS 处理前	排序	TOPSIS 处理后	排序
ANN MATH	70.63	1	93.31	1	70.52	1	91.61	1
J AM MATH SOC	65.12	2	89.68	2	59.80	2	84.41	4
COMMUN PUR APPL MATH	57.49	3	88.54	3	58.48	3	85.73	3
INVENT MATH	53.35	4	87.25	4	55.94	4	85.86	2
J DIFFER EQUATIONS	49.19	5	82.58	5	52.76	5	79.07	6
DUKE MATH J	41.28	11	81.75	6	43.38	11	80.11	5
ACTA MATH-DJURSHOLM	46.45	8	81.53	7	47.85	9	77.36	7
ADV MATH	48.05	6	80.05	8	51.55	6	75.75	8
T AM MATH SOC	42.45	10	79.06	9	44.69	10	75.24	9
PUBL MATH-PARIS	46.13	9	78.00	10	49.06	8	72.01	14
J EUR MATH SOC	34.02	15	76.49	11	35.80	14	73.79	10
J FUNCT ANAL	37.14	13	76.33	12	40.51	12	72.71	12
FOUND COMPUT MATH	36.59	14	76.31	13	40.14	13	72.42	13
J REINE ANGEW MATH	31.03	16	75.32	14	29.07	25	72.96	11
J MATH ANAL APPL	47.22	7	74.98	15	51.03	7	68.92	22
MATH ANN	30.39	17	74.10	16	32.16	18	71.33	16
J MATH PURE APPL	28.42	22	73.83	17	29.13	24	71.67	15
ANN SCI ECOLE NORM S	30.26	19	73.73	18	31.03	21	71.00	17
COMMUN PART DIFF EQ	26.42	24	72.26	19	27.78	26	70.01	18

（续表 4）

期刊名称	线性加权处理前	排序	线性加权处理后	排序	*TOPSIS* 处理前	排序	*TOPSIS* 处理后	排序
GEOM FUNCT ANAL	28.50	21	71.86	20	31.82	19	68.96	21
CALC VAR PARTIAL DIF	26.39	26	71.84	21	27.71	27	69.42	19
NONLINEAR ANAL-THEOR	30.36	18	71.52	22	33.74	15	67.65	24
J DIFFER GEOM	25.74	27	71.51	23	25.29	32	69.05	20
COMPOS MATH	25.33	28	71.29	24	25.50	30	68.88	23
P LOND MATH SOC	23.59	32	69.75	25	24.19	33	67.20	25
INT MATH RES NOTICES	27.14	23	69.70	26	29.84	22	66.14	26
LINEAR ALGEBRA APPL	29.42	20	69.35	27	33.08	17	64.95	28
DISCRETE CONT DYN-A	23.47	33	68.49	28	23.79	34	65.39	27
AM J MATH	21.75	35	67.72	29	22.53	35	64.95	29
ANAL PDE	24.29	30	67.31	30	27.14	28	63.93	32

数据处理前后评价得分的区分度与数据分布信息如表 5 所示。在数据处理前，线性加权汇总得分及格的只有 2 种期刊，TOPSIS 评价得分及格的只有 1 种期刊，这无论如何是难以说服公众的，也难以得到期刊编辑的认同，毕竟是期刊影响力评价，不可能有这么大的差距。数据处理后，线性加权汇总的及格期刊数量达到 54 种，TOPSIS 评价得分及格的增加到 41 种，大大改善了及格率，提高了评价结果与期刊影响力的一致性。

表 5　数据处理前后评价结果区分度与数据分布

比较指标	线性加权原始数据	线性加权处理数据	TOPSIS 原始数据	TOPSIS 处理数据
中位数/极大值	0.133	0.510	0.125	0.480
离散系数	0.812	0.271	0.906	0.289
HHI 系数	0.005 6	0.003 7	0.006 2	0.003 7
偏度 S	2.705	0.577	2.403	0.652
峰度 K	11.737	3.327	9.645	3.364
JB 检验	1 293.634	17.618	823.972	22.460
概率 p	0.000	0.000	0.000	0.000
及格率	2/294	54/294	1/294	41/294

从区分度看，数据处理前线性加权汇总评价得分的中位数极大值比为 0.133，TOPSIS 评价得分的中位数极大值比为 0.125，均比较低，即满分 100 分的情况下，至少一半期刊得分在 12.5 分以下，这也是不太符合期刊影响力实际的。经过数据处理后，线性加权汇总得分的中位数极大值比为 0.510，TOPSIS 评价得分的中位数极大值比为 0.480，这种数据分布已经比较均匀了。从离散系数看，指标数据处理前线性加权汇总得分的离散系数为 0.812，处理后降到了 0.271；指标数据处理前 TOPSIS 评价得分离散系数为 0.906，处理后降到了 0.289，均有大幅度改善。集中度郝氏指数也一样，数据处理前线性加权汇总得分 HHI 指数为 0.005 6，处理后降到 0.003 7；数据处理前 TOPSIS 评价的 HHI 指数为 0.006 2，处理后降到 0.003 7，改善较大。

从数据分布看，虽然无论数据处理前后评价结果均不服从正态分布，但是线性加权汇总 JB 检验值从 1 293.634 降到 17.618，TOPSIS 评价 JB 检验值从 823.972 降到 22.460，数据处理后使得评价结果更加接近正态分布。偏度虽然还是右偏，但也得到了大幅改善，线性加权评价从 2.705 下降到 0.577，TOPSIS 评价从 2.403 下降到 0.652。

5　研究结论

5.1　科技评价区分度低的根源是评价指标值与评价属性背离

本文分析了在科技评价尤其是文献计量指标中存在区分度与数据分布异常的问题，即低分数据过分拥挤，而高分数据又过于分散，评价数据分布呈现偏态分布。在此基础上如果进行多属性评价，那么评价结果也会存在类似问题。从而导致评价得分难以得到公认，评价公信力下降。产生这些问题的根本原因是评价指标值与评价属性背离，比如影响因子、总被引频次等不能客观反映影响力。只有那些真正反映学术创新的指标才不会出现评价指标值与评价属性背离，但目前这些指标还较少，数据获取困难。

5.2　引文指标更容易出现评价指标值与评价属性背离问题

文献计量指标的引文相关指标更容易出现评价指标值与评价属性的背离现象，尤其是影响力指标，如总被引频次、影响因子、他引影响因子、h 指数、5 年影响因子、即年指标、特征因子、标准特征因子等。对于其他时效性指标、期刊来源指标、编辑出版指标等需要进一步分析。

5.3 可以从多角度判定评价指标值与评价属性背离问题

判定评价指标值是否准确反映评价属性的方法较多,可以从多角度进行。第一是评价属性自身的内涵,其是否真的具备"一览众山小"的特征,比如重大原始创新、重大基础创新等,因为一般只有具备这些特点的评价指标才有可能出现数据分布的严重有偏现象。其次是从及格率、离散系数、中位数极大值比、集中度指数 HHI 等方面综合进行判断。

5.4 采用对数中位数标准化可以大幅降低评价指标值与评价属性背离问题

为了降低评价指标值与评价属性的背离现象,本文提出了对数中位数标准化方法,其原理是首先对评价指标取自然对数,从而缩小极大值与极小值的差距,并使得数据分布更加接近正态分布。然后再进一步进行标准化处理,对于中位数位于极差一半之前的评价指标,直接采用传统方法进行标准化;对于中位数位于极差一半之后的评价指标,需要加上该相对差值然后再做标准化。实证研究结果表明,采用对数中位数标准化方法处理后,评价指标和评价值的中位数极大值比得到有效改善,离散系数和集中度更加均匀,数据分布更加接近正态分布,使得评价指标值更加能够表达评价属性值。

5.5 建议评价时采用对数中位数标准化结果

在科技评价中,如果出现评价指标与数据分布异常现象,在认真分析后,如果确认存在评价值与指标属性背离现象,建议采用对数标准化后的结果进行单个指标评价与多属性评价,以便真实反映评价对象的属性情况与总体情况。

参考文献:

[1] VINKLER P. Introducing the current contribution index for characterizing the recent, relevant impact of journals [J]. Scientometrics,2008,79(2) : 409 －420.

[2] SEGLEN P O. The skewness of science[J]. Journal of the American Society for Information Science,1992,43(9) : 628－638.

[3] BORNMANN L,MUTZ R,NEUHAUS C,et al. Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results[J]. Ethics in science and environmental politics(ESEP),2008,8(1) : 93－102.

[4] ADLER R, EWING J,TAYLOR P. Citation statistics[J]. Statistical science, 2009,24(1): 1－26 .

[5] VAN RAAN A F J. Measurement of central aspects of scientific research: performance, interdisciplinarity, structure[J]. Measurement,2005, 3(1):1－19.

[6] 苏为华. 多指标综合评价理论与方法问题研究[D]. 厦门:厦门大学,2002.

[7] DAVIS P M. Eigenfactor: does the principle of repeated improvement result in better estimates than raw citation counts[J]. Journal of the American Society for Information Science and Technology, 2008,59(13): 2186－2188.

[8] HIRSCH J. An index to quantify an individuals scientific research output[J]. Proceedings of the National Academy of Sciences,2005, 102 (46):16569－16572.

[9] GLÄNZEL W. On the h-index - a mathematical approach to a new measure of publication activity and citation impact[J]. Scientometrics, 2006, 67(2):315－321.

[10] 俞立平,潘云涛,武夷山. 高校社科类学术期刊分级管理存在的问题研究[J]. 情报杂志,2011(4):1－4.

[11] 文东茅,鲍旭明,傅攸. 等级赋分对高考区分度的影响——对浙江"九校联考"数据的模拟分析[J]. 中国高教研究,2015(6): 17－21,72.

[12] RUANE F, TOL R. Rational (successive) h-indices: an application to economics in the Republic of Ireland[J]. Scientometrics, 2008,75(2):395－405.

[13] 叶鹰. 对数 f 指数及其评价学意义[J]. 情报科学,2009,27(7):965－968.

[14] 付鑫金,方曙,许海云. 高校网络学术影响力实证研究[J]. 图书情报工作,2013,57(8):86－90.

[15] 许新军. 期刊 hc 指数与 hd 指数的比较分析[J]. 情报杂志, 2015,34(4):59.

[16] 刘学之,杨泽宇,沈凤武,等. 基于 S 型曲线的指标非线性标准化研究[J]. 统计与信息论坛,2018,33(2):17－21.

[17] 郭亚军,马凤妹,董庆兴. 无量纲化方法对拉开档次法的影响分析[J]. 管理科学学报,2011(5):19－28.

[18] 俞立平,潘云涛,武夷山. TOPSIS 在期刊评价中的应用及在高次幂下的推广[J]. 统计研究,2012(12):96－101.

[19] 俞欣辰,潘有能. 学术期刊评价指标与评价结果的数据分布关系研究[J]. 情报杂志,2015,34(9):102－106.

[20] 俞立平,刘爱军. 指标数据分布与内部差距对学术期刊评价的影响——以 JCR 数学期刊为例[J]. 图书情报工作,2014,58(21):105－110.

[21] 詹敏,廖志高,徐玖平. 线性无量纲化方法比较研究[J]. 统计与信息论坛,2016(12):17－22

[22] UNDP. Human development report technical notes 2014[R/OL]. [2018－06－01]. http://hdr. undp. org/sites/default/files/hdr14_technical_notes. pdf.

[23] 苏为华. 对数型功效系数法初探[J]. 统计研究,1993(3):63－66.

[24] 封婷. 综合评价中一种凹性指数型功效函数[J]. 统计与信息论坛,2016,31(7):16－21.

[25] 王连芬,张少杰. 产业竞争力的测度指标体系设计[J]. 统计与决策,2008(10):49－50.

[26] HIRSCHMAN A O. The political economy of import - substituting

industrialization in Latin America[J]. The quarterly journal of economics, 1968(1):1 - 32.

作者贡献说明:

周娟美:标准化方法设计;
郭强华:研究方案设计;
王作功:数据分析、清洗、处理;
俞立平:论文总体设计、撰写。

Research on the Deviation Cause and Correction of Technology Evaluation Index Value and Evaluation Attribute

Zhou Juanmei[1] Guo Qianghua[2] Wang Zuogong[3] Yu Liping[4]
[1] School of Economics and Management, North University of China, Taiyuan 030051
[2] Economics School, Yunnan University, Kunming 650504
[3] Guizhou University of Finance and Economics, Institute of Green Development Strategy in Western China, Guiyang 550025
[4] School of Management and E-business, Zhejiang Gongshang University, Hangzhou 310018

Abstract: [**Purpose/significance**] This paper analyzes the deep-level reasons for the abnormality in the index distinction degree and the bias of data distribution in the science and technology evaluation and believes that this is a deviation between the evaluation index value and the evaluation attribute essentially. That is, the evaluation index value can not reflect the essence of the evaluation attribute well. [**Method/process**] This paper proposes a new method to reduce the deviation between evaluation index value and evaluation attribute which is logarithmic median standardization, and takes JCR2016 mathematics journal for an example to conduct an empirical analysis. [**Result/conclusion**] The results show that citation indexs are more likely to show deviations between evaluation index value and evaluation attribute. The deviation between evaluation index value and evaluation attribute can be determined from multiple perspectives, such as index connotation analysis, pass rate, dispersion coefficient, median maximum ratio, concentration index HHI, etc. The use of logarithmic median standardization can greatly reduce the deviation of evaluation index value and evaluation attribute.

Keywords: distinction degree data distribution science and technology evaluation influence index logarithmic median normalization